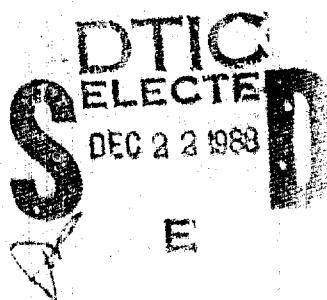AD-A202 066

RSRE
MEMORANDUM No. 4194

ROYAL SIGNALS & RADAR
ESTABLISHMENT

SPEECH AND LANGUAGE TECHNOLOGY:
A UK STRATEGY (SALTUS-2000)

Author: R K Moore

PROCUREMENT EXECUTIVE,
MINISTRY OF DEFENCE,
P C R E MALVERN,
WORCS.

88 12 22 037

- A -

ROYAL SIGNALS AND RADAR ESTABLISHMENT

Memorandum 4194

TITLE: SPEECH AND LANGUAGE TECHNOLOGY: A UK STRATEGY (SALTUS-2000)

AUTHOR: R. K. Moore

DATE: July 1988

SUMMARY

This memorandum contains the text of a discussion document prepared by the author in his position as Chairman of the UK Institute of Acoustics Speech Group. The document introduces 'SALTUS-2000': a strategy for UK research into speech and language technology. The main recommendations are: the establishment of a single competitive-collaborative research programme involving all of the main UK research groups, a public domain task portfolio, parallel civil and defence applications, widely available recordings and databases, agreed performance targets and prescribed assessment procedures.

The SALTUS-2000 proposals were presented to the UK speech research community in an address given at the Federation of the Acoustical Societies of Europe (FASE) conference on Speech (SPEECH-88) held in Edinburgh over the period 23-26 August 1988.

Crown copyright: Great Britain, (ICP)

CONTENTS

1. INTRODUCTION

Thanks mainly to the Alvey programme, Speech Technology research in the UK has received unprecedented levels of funding over the past three years. The size of the research community is significantly larger than it was prior to 1983 and considerable progress has been made in establishing alternative groups with different approaches to the problems of applying speech technology in the human-computer interface.

1988 is a critical year for UK speech technology research; the foundations of a strong research community have been laid but many Alvey (and ESPRIT I) projects are coming to an end. On the other hand ESPRIT II and ESPRIT BRA projects are in the process of being selected for funding and a new IED programme has been proposed.

It is an appropriate time to consider the wider strategic position for UK speech technology research and to provide a backdrop for the IED programme and any other nationally funded programmes which might emerge in the next decade.

Such a consideration is especially important if the UK is achieve a significant market position in this technology given the advances that are being made overseas (particularly in the USA).

2. BACKGROUND

The task of large vocabulary speaker independent automatic speech recognition has long been a target for research groups around the world. Since the early 1960s there have been two competing approaches: one based on the explicit use of speech knowledge in the form of phonetic rules and the other based on speech pattern matching. For the last ten years the latter approach has received considerable attention and has now developed into a sophisticated statistical methodology for modelling speech patterns. The approach is based on the mathematical technique of hidden Markov modelling (HMM) and work has reached the stage where explicit phonetically motivated sub-word units are invoked. Using this approach it is possible to demonstrate in near real-time a speaker independent word recognition accuracy of around 95% on a vocabulary of 1000 words with a grammatical perplexity of about 60 [1].

Such performance has been demonstrated by several research groups notably in the USA by collaborating laboratories within the DARPA Strategic Computing Speech and Natural Language Programme.

Despite considerable experience in the UK in these areas, no single research group has yet demonstrated an equivalent level of performance. The UK is currently lagging behind the rest of the world - not in ideas, but in an ability to put the ideas into practice.

The reason for this is the lack of any real collaboration between the different industrial, academic and government groups in the UK. This is surprising since the UK speech research community is small and tightly knit and considerable

enthusiasm exists for collaboration. In fact the Speech Technology Assessment Group (STAG) was formed in December 1983 in response to a bottom-up push from the grass roots of the speech community who perceived the importance of agreed testing protocols [2]. This initiative was followed closely in January 1984 when the community held a workshop to define the Alvey Speech Technology Strategy based on a market driven pull [3].

Unfortunately, the resulting ALVEY programme (including the Major Demonstrator) did not pull the community together until the very end; the funded projects have achieved useful results within themselves, but UK Ltd. has not benefitted greatly. Nevertheless the programme did succeed in training-up staff who now have the necessary experience and facilities to tackle the original objectives.

The community now has a second chance, particularly since the IED have announced a new programme of funded research in this area [4]. However, this time it is important that a national strategy is defined to respond to the challenge from overseas - a strategy that builds on the proposals in the IED Systems Architectures document, pulls together the different speech technology research groups in the UK and links the speech and natural language research communities.

SALTUS-2000 is a proposal for such a national strategy - a background against which the proposed IED programme and any other funding initiatives can be addressed.

3. AIMS OF SALTUS-2000

In the UK it is extremely unlikely that there will ever be a 100% funded programme of a size which approaches that of the current DARPA activity. This means that if the UK is serious about speech and language technology research, it cannot afford to dissipate effort by attempting to cover all angles at an early stage nor can it tolerate an excessive duplication of effort.

It is also interesting to note that the DARPA successes have been partially mediated by encouraging a competitive element within the overall collaboration with specific laboratories responsible for supplying data and others responsible for defining and administering progress evaluation procedures.

Therefore, for the UK to make significant progress with the resources available it is recommended that there should be a single collaborative programme (*SALTUS-2000*) in which each of the main UK research groups plays a part - regardless of the availability and level of funds. To support this it is also recommended that there should be a public domain task portfolio with parallel civil and defence application specifications and that appropriate recordings and databases should be made widely available. Furthermore the programme should include prescribed training and testing procedures targetted specifically at the programme goals.

It is also recommended that a competitive element within the programme would act as a significant pull on the individual research workers.

Within the competitive collaboration there should be coordinated research

objectives and specific goals which are directed towards providing the UK with the necessary algorithmic knowledge such that the UK will have sufficient expertise to attempt complete coverage of all application areas by the year 2000.

This can be mediated by a common module approach using agreed interface protocols between carefully defined levels of a speech and language processing system.

4. SALTUS-2000 TECHNICAL PROGRAMME

The key requirements for improved speech and language technology systems are: increased robustness, more natural speaking style, larger vocabularies, wider speaker population, minimum enrolment times, habitable applications interface, realistic environment, modular design, real-time operation and minimised product cost.

Systems may support interactive applications (such as voice control, voice data entry or voice data retrieval) or non-interactive applications (such as transcription, translation or vocoding). Other applications include wordspotting and speaker/accent/language identification and verification.

Interactive applications necessarily involve dialogue and therefore natural language processing. Many applications require speech synthesis as well as speech recognition.

The main challenge of *SALTUS-2000* is to identify two or three tasks in which all of the main research groups in the UK are prepared to take an interest.

The specific tasks addressed under the *SALTUS-2000* programme should thus be the subject of discussion within the community. However it is strongly recommended that in the short-term (one year) a serious attempt should be made to achieve a demonstration in the UK of a speech recognition capability which matches that currently available in the USA and in other countries.

In the long-term, the most important goal is to work towards systems with

unlimited vocabularies, particularly for transcription and for very low bit rate vocoding. Such a long-term goal would exert a strong pull on various key areas of speech and language technology research, for example a transcription system must be able to invoke the spelling of a word that has never been heard before, an interactive system must hypothesise the meaning of unknown words and be able to synthesise them and a vocoding system will require recognition and re-synthesis techniques which retain the characteristics of the original speaker.

Work directed towards these long-term goals will have significant short and medium-term spinoff (e.g. small vocabulary, low cost products) if they are addressed within the competitive collaborative environment proposed under *SALTUS-2000*.

As a result of these proposals it will be necessary to establish a new national archive of spoken English, and all such material should be made available with an appropriate level of annotation.

It is also recommended that an efficient way of approaching the issue of integrating natural language and speech is to make good use of full-speed simulations such as the Palantype shorthand machine transcription system. Much basic research in the middle ground between these two disciplines can be achieved by these means.

There will also be a continuing requirement for adequate support for basic research in human speech production, perception and dialogue.

5. IMPLICATIONS FOR PROPOSALS UNDER THE IED PROGRAMME

The proposed IED programme presents a unique opportunity for the broad aims of *SALTUS-2000* to be realised. Since *SALTUS-2000* derives from the IED strategy itself, it is recommended that the IED programme be used to create the open competitive and collaborative research environment proposed in this document.

This means that there is an opportunity to coordinate other projects not funded under the IED programme with relevant IED funded enabling projects and with the proposed IED major demonstrator(s).

6. IMPLICATIONS FOR OTHER FUNDING AGENCIES

SALTUS-2000 will have maximum impact if it is taken into account by the new SERC/DTI joint committee(s) responsible for funding speech and language technology research work and by other funding agencies such as the MoD.

7. SUMMARY

SALTUS-2000 is a strategy for UK research in speech and language technology which is intended to establish a competitive collaborative environment suitable for the growth of a dominant UK position in this field by the year 2000.

The main recommendations are: the establishment of a single competitive collaborative research programme involving all of the main UK research groups, a public domain task portfolio, parallel civil and defence applications, widely available recordings and databases, agreed performance targets and prescribed assessment procedures.

REFERENCES

[1] Kai-Fu Lee. *Large Vocabulary Speaker-Independent Continuous Speech Recognition: The SPHINX System.* Ph.D. Thesis, Computer Science Department, Carnegie-Mellon University, Pittsburgh, USA, April 1988.

[2] *Speech Input/Output Assessment.* Proc. Inst. Acoustics Speech Group Workshop, London, 1 December 1983.

[3] *Speech Technology in the Alvey Programme.* Proc. Alvey Speech Technology Workshop, Coventry, 23-24 January 1984.

[4] *Systems Architecture: A Strategy for Research in Parallel Architecture, Distributed Systems, Vision and Speech Technologies.* Report by the DTI Information Engineering Directorate, June 1988.

DOCUMENT CONTROL SHEET

Overall security classification of sheet ...UNCLASSIFIED.. ....

(As far as possible this sheet should contain only unclassified information. If it is necessary to enter
classified information, the box concerned must be marked to indicate the classification eg (R) (C) or (S))

1. DRIC Reference (if known)	2. Originator's Reference Memorandum 4194	3. Agency Reference	4. Report Security Classification Unclassified
5. Originator's Code (if known) 7784000	6. Originator (Corporate Author) Name and Location Royal Signals and Radar Establishment St Andrews Road, Malvern, Worcestershire WR14 3PS		
5a. Sponsoring Agency's Code (if known)	6a. Sponsoring Agency (Contract Authority) Name and Location		

7. Title   SPEECH AND LANGUAGE TECHNOLOGY: A UK STRATEGY (SALTUS-2000)

7a. Title in Foreign Language (in the case of translations)

7b. Presented at (for conference papers)   Title, place and date of conference

8. Author 1 Surname, initials Moore R K	9(a) Author 2	9(b) Authors 3,4...	10. Date 7.88	pp. ref. 10
11. Contract Number	12. Period	13. Project	14. Other Reference	

15. Distribution statement Unlimited

Descriptors (or keywords)   continue on separate piece of paper

Abstract

    This Memorandum contains the text of a discussion document prepared by the
author in his position as Chairman of the UK Institute of Acoustics Speech Group.
The document introduces 'SALTUS-2000': a strategy for UK research into speech and
language technology. The main recommendations are: the establishment of a single
competitive-collaborative research programme involving all of the main UK research
groups, a public domain task portfolio, parallel civil and defence applications,
widely available recordings and databases, agreed performance targets and
prescribed assessment procedures.
    The SALTUS-2000 proposals were presented to the UK speech research community
in an address given at the Federation of the Acoustical Societies of Europe (FASE)
Conference on Speech (SPEECH-88) held in Edinburgh over the period 23-26 August 88.

S80/48